# Resolving fine granularity toponyms: Evaluation of a disambiguation approach

Derungs, Curdin ; Palacio, Damien ; Purves, Ross S

# Resolving fine granularity toponyms: Evaluation of a disambiguation approach

C. Derungs, D. Palacio, R.S. Purves[1]

Department of Geography, University of Zurich, Winterthurstrasse 190, 8057 Zürich, Switzerland
{curdin.derungs, damien.palacio, ross.purves}@geo.uzh.ch

## 1. Introduction

Landscape descriptions in natural language, for instance from historic corpora, are a complementary source to empirical ethnographic work, for example to research exploring variation in the use of *basic levels* or *basic terms* within landscapes across localities (c.f. Mark and Turk 2003, Burenhult and Levinson 2008, Turk et al. 2011), on the condition that such descriptions can be linked to space. A key challenge in linking language to space is the detection and resolution of toponyms (Purves and Jones 2008). Central to toponym resolution is the identification of a single unambiguous referent for a given toponym, which requires that toponym referent ambiguity is resolved (c,f, Amitay et al. 2004), i.e. does the document refer to London, England or London, Ontario. Some common state of the art approaches to toponym disambiguation use *default rules*, such that the most prominent referent location is resolved (c.f. Purves et al. 2007), *population counts,* also reflecting the prominence of referent locations (c.f. Martins et al. 2010) and *geometric minimality*, assuming that the areal footprint of a document is to be minimised (c.f. Leidner 2004).

Leidner (2007) argued that toponym disambiguation had focused on populated places, since such locations are important for a variety of applications (e.g. local search or news mapping). However, if we wish to resolve toponyms with a fine spatial granularity, such as those typically used to reference mountains, hills, fields or hamlets in natural landscape descriptions, state of the art disambiguation approaches must be adapted to work independently from *a priori* toponym knowledge that is usually attached to populated places and commonly found in gazetteers (Hill 2006).

We present an approach for toponym disambiguation, working independently from a priori toponym knowledge. We evaluate its performance over a baseline disambiguation technique on an extensive corpus consisting of 150 years of Swiss alpine literature (Volk et al. 2009). Toponym knowledge is gathered from geomorphometric characteristics at locations of toponyms. This reflects the strong relation between toponyms and topography, since toponyms are used to name geographic objects that are attached to the earth's surface (Smith and Mark 2003) and therefore can be hypothesised to be bound to its characteristics.

We show that in a user-centered evaluation with spatial queries (i.e. classifying articles contained in the alpine corpus as being relevant or not for certain spatial extents) our approach to disambiguation, using geomorphometric information, significantly outperforms baseline disambiguation (27% improvement).

## 2. Geomorphometric Toponym Disambiguation

Our approach combines, firstly the concept of geometric minimality, reflecting the notion of spatial autocorrelation commonly used in toponym disambiguation (Leidner 2004) and, secondly, geomorphometric characteristics at toponym locations. Geomorphometric characteristics form toponym knowledge which is usually missing when working with landscape descriptions of fine spatial granularity. In two case studies we showed, firstly, that

using geomorphometric characteristics in toponym disambiguation of a corpus describing one type of toponym, namely *Hochmoor* (raised bog), increased accuracy (58% vs. 23%) (Derungs et al. 2011), and, secondly, that such geomorphometric characteristics can be gathered and structured in a way that toponyms of very different types can be distinguished (Derungs and Purves 2012). In both case studies we gathered topographic characteristics from geomorphometric classifications of elevation models at a range of scales (Figure 1).

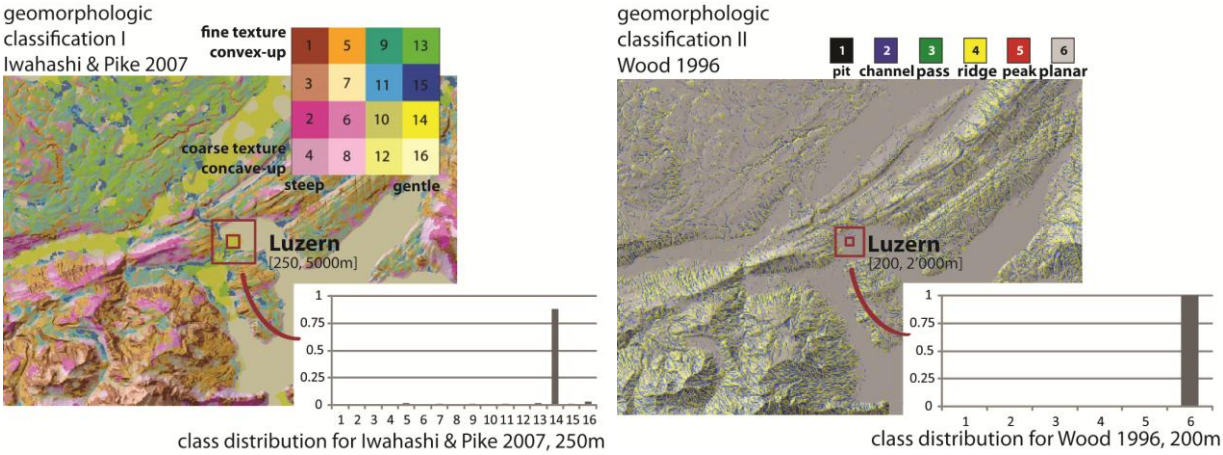


Figure 1. Geomorphometric classification of an elevation model according to Iwahashi and Pike (2007, left) and Wood (1996, right).

In this paper we demonstrate our approach to toponym disambiguation on a corpus containing all types of toponyms (which usually is the case). Therefore we assume that for a word, having the same wording as a toponym (*potential toponym*) to be resolved as a toponym and to be linked to a referent location, it needs to fit either to the geomorphologic or geometric context of a particular paragraph in the article.

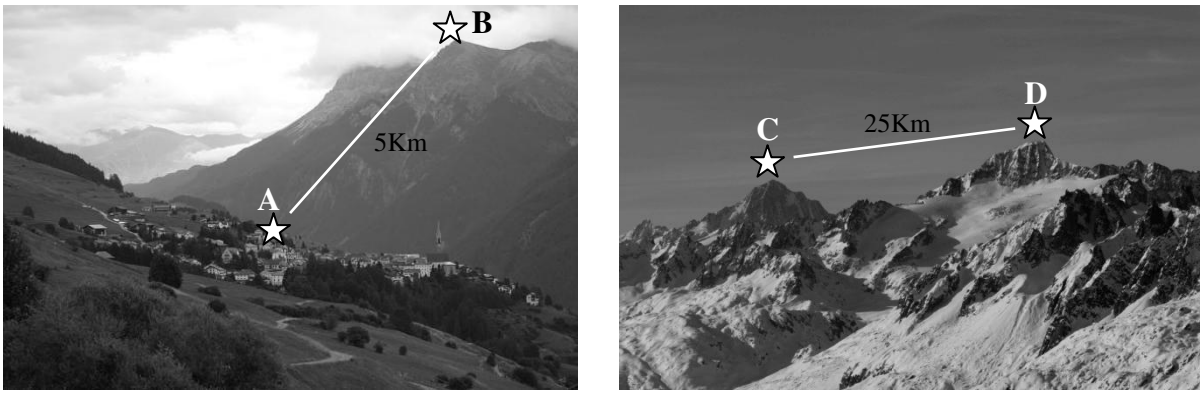


Figure 2. Two photographs visualising good fit between toponyms of either geometry (left, A and B) or geomorphometry (right, C and D).

In Figure 2 two prototypes of such fits are demonstrated: The geometric fit between the village of Sent (A) and the mountain Piz Lischana (B) is good because the two locations are proximate in Euclidean space. The geomorphometric fit between the mountains Finsteraarhorn (C) and Gallenstock (D) is good because both locations refer to mountains and therefore share geomorphometric characteristics. The value of 'fit' is calculated with respect to a varying set of locations gathered from all potential toponyms with all referent locations within the same text document.

The geometric and geomorphometric thresholds of a text are first gathered from a set of potential toponyms that fulfil two conditions: a) only have one referent location and b) show below average web-counts as measured using the Yahoo! Search BOSS API (c.f. Pasley et al. 2008). Web-counts are incorporated as an index of proneness of a potential toponym to semantic ambiguity. The geometric and geomorphometric scope of a text document is iteratively updated with each next potential toponym that is tested for fit.

## 3. Case study: Evaluation with Text + Berg test collection

### 3.1 Text + Berg: A historic corpus of Swiss Alpine Literature

One year after the Swiss Alpine Club was founded, in 1864, the first yearbook describing its activities was published. By 2009 the Text + Berg corpus consisted of 134 yearbooks, each with 300 to 600 pages from an average of 80 articles, that are digitised and part-of-speech tagged (Volk et al. 2009). The topical focus of Text + Berg ranges from classical and modern mountaineering, and includes a wide range of literary styles.

### 3.2 Evaluation protocol

We evaluated the improvement our paper brought to a geographic information retrieval task in comparison to a baseline approach. To perform such an evaluation, we require a test collection, such as is typical in information retrieval exercises such as TREC (Voorhees and Harman 2005) comprised of: a set of queries representing the user's need, a corpus, in our case Text + Berg, and query relevance judgment obtained by asking users to determine if results are relevant to the query.

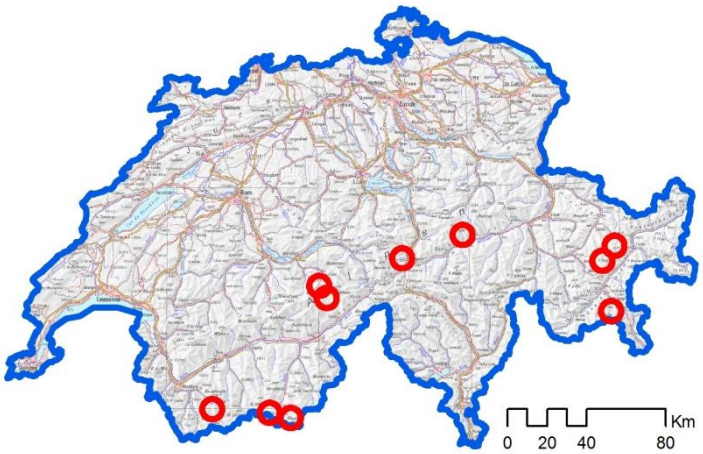


Figure 3. 10 spatial queries.

We choose 10 spatial queries from Swiss mountain regions (see Figure 3), well covered by articles from Text + Berg. We submitted the 10 queries to 2 approaches: a simple one that randomly selects a referent location in case of ambiguity (a typical *baseline* approach where no other knowledge is available, c.f. Clough 2005), and, the *geometric and geomorphometric disambiguation (GGD)*, described above. Relevance ranking of articles incorporates the number of spatial references inside and outside the spatial query of each article, and for *GGD* also the fit of each referent is taken into account. From both approaches we selected the top 5-6 ranked articles and merged them to form a list containing unique articles (all lists contained at least 9 articles). The relevance of the articles for each query was judged by 5 test users (with a total of 12 participants). Since annotating articles for relevance for a spatial extent is highly dependent on local knowledge (c.f. Purves et al. 2007), our participants had

knowledge of the Swiss alps (e.g. through physical geography or mountaineering). In addition, we provided participants with detailed topographic maps (1:50000 and 1:25000).

## 4. Result

As shown in Figure 4, an average of 82% of the top 5-6 articles gathered from each of the 10 spatial queries, using *GGD*, were judged to be relevant. This is practically and statistically significantly higher than the 55% gained with the *baseline* approach (t-test: $p<0.05$).

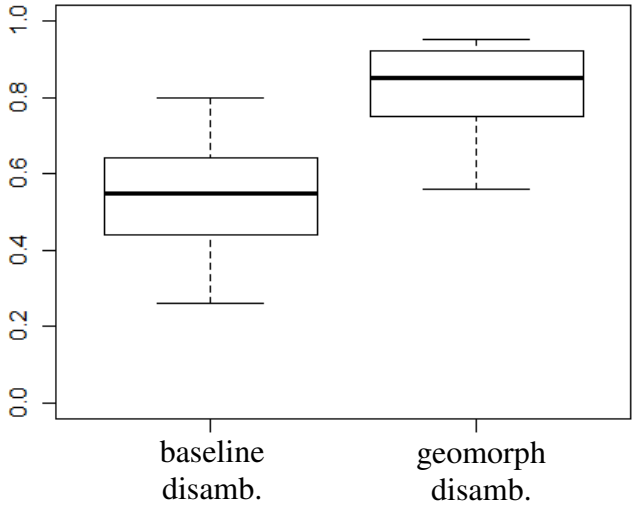


Figure 4. Relevance judgments for both disambiguation approaches.

Articles retrieved through geomorphometric disambiguation tend to be longer than those retrieved with the baseline, whereas baseline articles appear to be focussed on the topic when only titles are considered. This is due to the incorporation of 'fit' of spatial references when gathering articles through *GGD*. Articles thus need not be explicitly devoted to only one region. The ranking is good as soon as articles contain relevant spatial descriptions. In general participants seem to favour more extensive descriptions, however, for one particular query (Monte Rosa Region) the focus of titles appears to have strongly influenced relevance judgements (80% vs. 56% relevant articles).

## 5. Concluding Discussion

This paper is part of ongoing work with the aim of linking natural landscape descriptions with space. This is a first step to open up the extensive source of natural language descriptions to investigate how people conceptualise landscapes and its variation in space and time. However, this also implies working with a particular problematic set of toponyms, which lack *a priori* toponym knowledge due to their fine spatial granularity and type. We therefore incorporated geomorphometric characteristics in toponym disambiguation, as an addition to state of the art geometric minimality. An evaluation on the corpus of Text + Berg, a historic collection of Swiss alpine literature, showed results significantly improved compared to a baseline. In general, it would be interesting to use a more extensive set of queries, and also to compare against more sophisticated disambiguation approaches (even if they are not adjusted to fine spatial granularity toponyms), to confirm our results. However, local knowledge is a bottleneck in user-centered evaluations, particularly of fine spatial granularity information.

## Acknowledgements

The research reported in this paper was funded by the project FolkOnt supported by the Swiss National Science Foundation under contract 200021-126659.

## References

Amitay E, Har'El N, Sivan R, Soffer A, 2004, Web-a-Where: Geotagging Web content. In: Sanderson M, Järvelin K, Allan J, Bruza P (eds), Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, 25-29.

Burenhult N, Levinson SC, 2008, Language and landscape: A cross-linguistic perspective. *Language Sciences*, 30(2), 135-150.

Clough P, 2005, Extracting metadata for spatially-aware information retrieval on the Internet, In: Jones CB, Purves RS (eds) *Proceedings of the ACM Workshop on Geographic Information Retrieval*, Bremen, Germany, 25–30.

Derungs C, Purves RS, 2012, Measuring topographic similarity of toponyms. In: *Proceedings of the 15thAGILE International Conference on Geographic Information Science*, Avignon, France.

Derungs C, Purves RS and Waldvogel B, 2011, Toponym disambiguation of landscape features using geomorphometric characteristics. In: Cheng T (eds) *Proceedings of the 11th International Conference on GeoComputation*, London, UK, 106-110.

Hill L, 2006, *Georeferencing: The Geographic Associations of Information*, MIT Press, Cambirdge, UK.

Iwahashi J, Pike RJ, 2007, Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature, In: *Geomorphology*, 15(3):409-440.

Jones CB , Purves RS, 2008, geographical information retrieval. International Journal of Geographic Information Science, 22(3):219-228.

Leidner J, 2004, Which Sheffield is it? In: Sanderson M, Järvelin K, Allan J, Bruza P (eds), Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, 602-610.

Leidner J, 2007, Toponym Resolution in Text : Annotation, Evaluation and Applications of Spatial Grounding of Place Names, Universal Press, Florida, USA.

Mark DM, Turk A, 2003, Landscape categories in Yindjibarndi: Ontology, environment, and language. In: Kuhn W, Worboys MF, Timpf S, *Spatial information theory: Foundations of geographic information, Cosit 2003*, Ittingen, Switzerland, 28–45.

Martins B, Anastacio I, Calado P, 2010, A Machine Learning Approach for Resolving Place References in Text. Lecture Notes in Geoinformation and Cartography, 0:221-236.

Pasley R, Clough P, Purves RS, Twaroch FA, 2008, Mapping geographic coverage of the web. In: *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, New York, 1-9.

Purves RS, Cloub P, Jones CB, Arampatzis A, Bucher B, Finch D, Fu G, Joho H, Syed AK, Vaid S, Yang B, 2007, The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet. International Journal of Geographic Information Science, 21(7):717-745.

Smith B, Mark DM, 2003, Do Mountains exist? Towards an Ontology of landforms, In: *Environment and Planning B*, 30(3):411-428

Turk AG , Mark DM, Stea D, 2011, Ethnophysiography. In: Mark DM, Turk AG, Burenthult N, Stea D (eds), *Landscape in Language*, John Benjamins, Philadelphia, USA, 1–24.

Volk M, Bubenhofer N, Althaus A, Bangerter M, 2009, Classifying Named Entities in an Alpine Heritage Corpus. *Künstliche Intelligenz*, 4:40–43.

Voorhees EM, Harman DK, 2005, TREC: Experiment and Evaluation in Information Retrieval. MIT Press,Cambridge, USA.

Wood J, 1996, The geomorphological characterization of digital elevation models. PhD Thesis, University of Leicester, UK.